

Publications about Sensors and Sensor Networks

This page contains resources (peer-reviewed publications) that pertain to sensors, equipment, or data collection. For publications pertaining to quality control (statistics, data storage and versioning, etc.) please upload to the Quality Control Resources page.

- [Automated quality control methods for sensor data: a novel observatory approach, Taylor and Loescher, 2013](#), note: this PDF pertains to the NEON sites, both field sensor techniques and quality techniques

(abstract) National and international networks and observatories of terrestrial-based sensors are emerging rapidly. As such, there is demand for a standardized approach to data quality control, as well as interoperability of data among sensor networks. The National Ecological Observatory Network (NEON) has begun constructing their first terrestrial observing sites, with 60 locations expected to be distributed across the US by 2017. This will result in over 14 000 automated sensors recording more than > 100 Tb of data per year. These data are then used to create other datasets and subsequent “higher-level” data products. In anticipation of this challenge, an overall data quality assurance plan has been developed and the first suite of data quality control measures defined. This data-driven approach focuses on automated methods for defining a suite of plausibility test parameter thresholds. Specifically, these plausibility tests scrutinize the data range and variance of each measurement type by employing a suite of binary checks. The statistical basis for each of these tests is developed, and the methods for calculating test parameter thresholds are explored here. While these tests have been used elsewhere, we apply them in a novel approach by calculating their relevant test parameter thresholds. Finally, implementing automated quality control is demonstrated with preliminary data from a NEON prototype site.

- [Uncertainties in Selected River Water Quality Data, Rode, 2007](#)

(abstract) Monitoring of surface waters is primarily done to detect the status and trends in water quality and to identify whether observed trends arise from natural or anthropogenic causes. Empirical quality of river water quality data is rarely certain and knowledge of their uncertainties is essential to assess the reliability of water quality models and their predictions. The objective of this paper is to assess the uncertainties in selected river water quality data, i.e. suspended sediment, nitrogen fraction, phosphorus fraction, heavy metals and biological compounds. The methodology used to structure the uncertainty is based on the empirical quality of data and the sources of uncertainty in data (van Loon et al., 2005). A literature review was carried out including additional experimental data of the Elbe river. All data of compounds associated with suspended particulate matter have considerable higher sampling uncertainties than soluble concentrations. This is due to high variability within the cross section of a given river. This variability is positively correlated with total suspended particulate matter concentrations. Sampling location has also considerable effect on the representativeness of a water sample. These sampling uncertainties are highly site specific. The estimation of uncertainty in sampling can only be achieved by taking at least a proportion of samples in duplicates. Compared to sampling uncertainties, measurement and analytical uncertainties are much lower. Instrument quality can be stated well suited for field and laboratory situations for all considered constituents. Analytical errors can contribute considerably to the overall uncertainty of river water quality data.

- [Quantifying Uncertainty, Schofield, 2002](#)

(abstract) EPA's Great Lakes National Program Office (GLNPO) is leading one of the most extensive studies of a lake ecosystem ever undertaken. The Lake Michigan Mass Balance Study (LMMB Study) is a coordinated effort among state, federal, and academic scientists to monitor tributary and atmospheric pollutant loads, develop source inventories of toxic substances, and evaluate the fate and effects of these pollutants in Lake Michigan. A key objective of the LMMB Study is to construct a mass balance model for several important contaminants in the environment: PCBs, atrazine, mercury, and transnonachlor. The mathematical mass balance models will provide a state-of-the-art tool for evaluating management scenarios and options for control of toxics in Lake Michigan. At the outset of the LMMB Study, managers recognized that the data gathered and the model developed from the study would be used extensively by data users responsible for making environmental, economic, and policy decisions. Environmental

measurements are never true values and always contain some level of uncertainty. Decision makers, therefore, must recognize and be sufficiently comfortable with the uncertainty associated with data on which their decisions are based. The quality of data gathered in the LMMB was defined, controlled, and assessed through a variety of quality assurance (QA) activities, including QA program planning, development of QA project plans, implementation of a QA workgroup, training, data verification, and implementation of a standardized data reporting format. As part of this QA program, GLNPO has been developing quantitative assessments that define data quality at the data set level. GLNPO also is developing approaches to derive estimated concentration ranges (interval estimates) for specific field sample results (single study results) based on uncertainty. The interval estimates must be used with consideration to their derivation and the types of variability that are and are not included in the interval.

- [Know Thy Sensor, Wallis et al., 2007](#)

(abstract) For users to trust and interpret the data in scientific digital libraries, they must be able to assess the integrity of those data. Criteria for data integrity vary by context, by scientific problem, by individual, and a variety of other factors. This paper compares technical approaches to data integrity with scientific practices, as a case study in the Center for Embedded Networked Sensing (CENS). The goal of this research is to identify functional requirements for digital libraries of scientific data that will serve this community. Data sources include analysis of documents produced by the CENS data integrity group and interviews with science and technology researchers within CENS.

- [ICOADS Release, v. 2.1, Worley, 2005](#)

(abstract) The International Comprehensive Ocean–Atmosphere Data Set (ICOADS), release 2.1 (1784–2002), is the largest available set of in situ marine observations. Observations from ships include instrument measurements and visual estimates, and data from moored and drifting buoys are exclusively instrumental. The ICOADS collection is constructed from many diverse data sources, and made inhomogeneous by the changes in observing systems and recording practices used throughout the period of record, which is over two centuries. Nevertheless, it is a key reference data set that documents the long-term environmental state, provides input to a variety of critical climate and other research applications, and serves as a basis for many associated products and analyses. The observational database is augmented with higher level ICOADS data products. The observed data are synthesized to products by computing statistical summaries, on a monthly basis, for samples within 2° latitude \times 2° longitude and $1^\circ \times 1^\circ$ boxes beginning in 1800 and 1960 respectively. For each resolution the summaries are computed using two different data mixtures and quality control criteria. This partially controls and contrasts the effects of changing observing systems and accounts for periods with greater climate variability. The ICOADS observations and products are freely distributed worldwide. The standard ICOADS release is supplemented in several ways; additional summaries are produced using experimental quality control, additional observations are made available in advance of their formal blending into a release, and metadata that define recent ships' physical characteristics and instruments are available. Copyright © 2005 Royal Meteorological Society.

- [One Fish, Two Fish, Riddick, 2002](#)

(abstract) PA is conducting a National Study of Chemical Residues in Lake Fish Tissue. The study involves five analytical laboratories, multiple sampling teams from each of the 47 participating states, several tribes, all 10 EPA Regions and several EPA program offices, with input from other federal agencies. To fulfill study objectives, state and tribal sampling teams are voluntarily collecting pre-dator and bottom-dwelling fish from approximately 500 randomly selected lakes over a 4-year period. The fish will be analyzed for more than 300 pollutants. The long-term nature of the study, combined with the large number of participants, created several QA challenges: (1) controlling variability among sampling activities performed by different sampling teams from more than 50 organizations over a 4-year period; (2) controlling variability in lab processes over a 4-year period; (3) generating results that will meet the primary study objectives for use by OW statisticians; (4) generating results that will meet the undefined needs of more than 50 participating organizations; and (5) devising a system for evaluating and defining data quality and for reporting data quality assessments concurrently with the data to ensure that assessment efforts are streamlined and that assessments are consistent among organizations. This paper

describes the QA program employed for the study and presents an interim assessment of the program's effectiveness